Year: 2020

Spotting political social bots in Twitter: A use case of the 2019 Spanish general election

Pastor-Galindo, Javier ; Zago, Mattia ; Nespoli, Pantaleone ; López Bernal, Sergio ; Céldran, Alberto Huertas ; Pérez, Manuel Gil ; Ruipérez-Valiente, José A ; Pérez, Gregorio Martínez ; Mármol, Félix Gómez

DOI: https://doi.org/10.1109/TNSM.2020.3031573

Originally published at:
Pastor-Galindo, Javier; Zago, Mattia; Nespoli, Pantaleone; López Bernal, Sergio; Céldran, Alberto Huertas; Pérez, Manuel Gil; Ruipérez-Valiente, José A; Pérez, Gregorio Martínez; Mármol, Félix Gómez (2020). Spotting political social bots in Twitter: A use case of the 2019 Spanish general election. IEEE Transactions on Network and Service Management, 1(1):1-2.
DOI: https://doi.org/10.1109/TNSM.2020.3031573

Spotting political social bots in Twitter: A use case of the 2019 Spanish general election

Javier Pastor-Galindo*, Mattia Zago*, Pantaleone Nespoli*, Sergio López Bernal*, Alberto Huertas Celdrán†, Manuel Gil Pérez*, José A. Ruipérez-Valiente*, Gregorio Martínez Pérez*, Félix Gómez Mármol*
**Department of Information and Communications Engineering, University of Murcia*, 30100 Murcia, Spain
{javierpg, mattia.zago, pantaleone.nespoli, slopez, mgilperez, jruiperez, gregorio, felixgm}@um.es
†*Telecommunication Software & Systems Group, Waterford Institute of Technology*, Ireland
ahuertas@tssg.org

ABSTRACT

While social media has been proved as an exceptionally useful tool to interact with other people and massively and quickly spread helpful information, its great potential has been ill-intentionally leveraged as well to distort political elections and manipulate constituents. In the paper at hand, we analyzed the presence and behavior of social bots on Twitter in the context of the November 2019 Spanish general election. Throughout our study, we classified involved users as social bots or humans, and examined their interactions from a quantitative (i.e., amount of traffic generated and existing relations) and qualitative (i.e., user's political affinity and sentiment towards the most important parties) perspectives. Results demonstrated that a non-negligible amount of those bots actively participated in the election, supporting each of the five principal political parties.

I. INTRODUCTION

Social media have become one of the main channels to spread information worldwide at scale and their popularity renders them as one of the most impactful means of influencing public opinion [1], [2]. As stated by the 2019 Global Inventory of Organised Social Media Manipulation report [3] elaborated by the University of Oxford:

> *"Social media, which was once heralded as a force for freedom and democracy, has come under increasing scrutiny for its role in amplifying disinformation, inciting violence, and lowering levels of trust in media and democratic institutions."*

To this extent, one of the most powerful strategies to maximize the dissemination of a message that aims to deceive social media users consists on using social bots as amplifiers [4]. Those software-controlled social accounts are able to effectively mimic the normal behavior of human users while sneakily operating at a much higher rate and remaining obscure [5]. In particular, recent studies have disclosed how these coordinated armies are working to poison democratic elections in an orchestrated manner [6].

While social media enables the fast propagation of fake news or any other misleading information over the web, the so-called *political social bots* take care of amplifying their popularity to catch the eye of virtual communities and to create manually crafted viral trends [7].

They share a common characteristic: the abuse of automation tools to generate huge amounts of social media activity in order to support, or oppositely attack, political figures following their agenda with personal interests. Alarmingly, bots are progressively becoming more sophisticated thanks also to the advances in Artificial Intelligence [8]. Slowly, year by year, bots can build more realistic social media behaviors and produce credible content with human-like temporal patterns [9], while creating a network of interconnections to spread the forged information further [10].

In this regard, social bots represent a growing phenomenon (see Figure 1) which aims at jeopardizing modern democracies by distorting reality and manipulating constituents [11]. These malicious operations are often referred to as astroturfing or Twitter bombs, which fake the appearance of organic grassroot participation while being secretly orchestrated and funded [12]. In such a scenario, it is clear that this threat is more present and real than ever, potentially affecting millions of users that are absolutely unaware of these malicious activities which may undermine worldwide democracies [13].

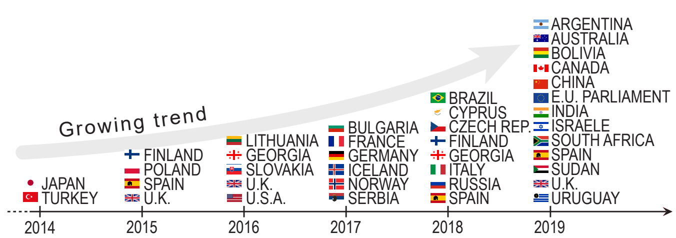


Fig. 1: Recent political and administrative elections with social bots participation.

Given the above-mentioned non-negligible threats and dangers to our democracies and modern societies, there is an imperative and urgent need to develop innovative solutions to *i)* defeat evildoers aiming at subverting political elections by leveraging armies of social bots and to *ii)* safeguard defenseless citizens from ill-intentioned manipulation. In this sense, Open Source Intelligence (OSINT) [14] becomes a promising paradigm with which to perform an in-depth analysis of publicly accessible sources, such as social networks, and tackle social media manipulation.

In this work, we analyzed the presence and behavior of social bots on Twitter in the context of the November 2019 Spanish general election. In particular, we spotted those social bots participating during the election campaign in Twitter, measuring their activity, conducting a temporal analysis of their behavior, studying their interactions, and, most importantly, assessing their sentiment towards the top five political parties in the election.

To this end, the paper at hand is structured as follows: *i)* Section II will provide both the political context (Section II-A) and a review of the literature (Section II-B); *ii)* Section III will provide the required information regarding the followed methodology and the mathematical background; *iii)* Section IV will present the results of the analysis which are later commented in *iv)* the discussion section (V); finally, *v)* Section VI will conclude and examine potential research lines.

II. Background

For the development and understanding of the case study it is necessary to have a basic notion of the Spanish political context, discussed in Section II-A. Meanwhile, we comment on how this type of problems have been tackled in Section II-B.

A. Political Context

On November 10th, 2019, the Spanish general election was held, where five dominant political parties (among several others) participated, namely: United We Can (*Unidas Podemos*, UP, left-wing to far-left), the Spanish Socialist Worker's Party (*Partido Socialista Obrero Español*, PSOE, centre-left), Citizens (*Ciudadanos*, Cs, centre to centre-right), the People's Party (*Partido Popular*, PP, centre-right to right-wing) and VOX (VOX, right-wing to far-right).

Such fragmentation made it difficult for any other formation or coalition to achieve an electoral majority. In fact, PSOE won the election and, at the time of writing this article, PSOE and UP govern Spain in coalition.

Throughout the time window analyzed in this work, a number of remarkable events should be highlighted:

- *October 10th, 2019* – Santiago Abascal (VOX's political leader) participated in the live TV show "El Hormiguero"
- *October 19-20th, 2019* – Riots in Catalonia
- *October 24th, 2019* – Exhumation of Francisco Franco
- *November 4th, 2019* – Electoral debate on national TV
- *November 10th, 2019* – General election day

Finally, Figure 2 reports the parliament composition as a result of the general election, note that only the five main parties are highlighted.

B. Related Works

As already mentioned, social bots persist as a severe threat against modern digital democracies, attracting the attention of both academia and industry. Several researchers worldwide studied such alarming phenomenon, aiming at shedding light and proposing effective solutions.

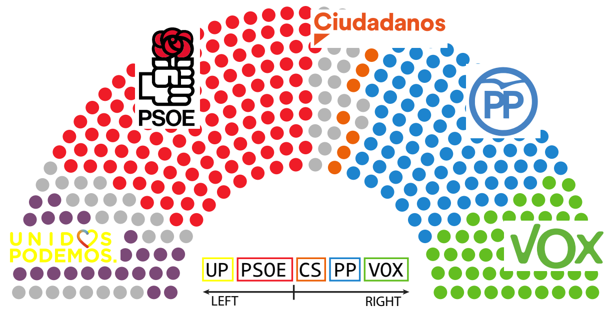


Fig. 2: Parliament composition and parties' disposition after the 2019 Spanish general election.

In [4], the authors compared the communication patterns among Twitter users during riot events. In particular, they studied the emotional level of the messages, with a particular focus on the direct messages generated by bot accounts. Remarkably, bots convey emotions that are comparable to those conveyed by human accounts. Such a strategy is indeed used by the bots to remain unveiled and, thus, attract humans that are encouraged to interact with automatic-generated content. Similarly, in [15], the relative importance and persistence of social bots were analyzed during 3 crucial events, namely: 2016 US presidential elections, the ongoing Ukrainian-Russian conflict, and the Turkish censorship implemented by the government. More specifically, the primary outcomes of the study showed that bots attempted to initiate contact with users at an extremely higher rate than human users. Through the application of social network analysis centrality measurements, authors found out that social bots, while representing less than 1% of the total user population of the dataset, displayed an incredible level of structural network influence, with bot influencers capable of effortlessly captivating human ones.

Alarmingly, the social bots activity emphasizes its effectiveness during the political election campaigns, trying to deceive worldwide citizens toward forged viral trends. In [8], an overview of the bots' activities during crucial political elections (i.e., 2016 US presidential, 2017 French presidential, and 2018 US midterm) is presented, highlighting on some significant peculiarities. More specifically, the author uncovered massive participation of such accounts, which showed human-like behavior while performing social interactions, and, dangerously, the capability of adapting themselves to remain obscure to the Twitter detection modules. As suggested in other relevant articles, the author claimed that a black-market for political bots might exist. Recently, sophisticated detection methodologies have been proposed to win this arm race against the bot collusion as an ultimate goal. In particular, in [16], authors proposed an adversarial approach to detect evolving bots. To this extent, authors synthetically modify existing bots leveraging genetic algorithm, and demonstrated the effectiveness of their proposal. Nonetheless, one could argue that the generality of such detection methodology is questionable due to the crafted nature of the algorithmically generated bots. Furthermore, an approach based on Inverse Reinforcement Learning (IRL) aiming at capturing bot behavior and thus identify bot accounts was presented in [17]. By leveraging

2016 US presidential election data, authors were capable of correctly classifying the majority of social bots by relying on the flow of online activity within the social platform. However it is realistic to claim that more perspectives need to be considered in order to defeat the bots' army, such as the social network structure and the posted content, just to cite some examples.

To the best of our knowledge, none of the presented research works performed an in-depth analysis of the 2019 Spanish general election. Specifically, we believe that an effort is more than necessary to clarify further the role of those forged accounts in a crucial scenario for the European geopolitical scene. Additionally, by studying the social interaction and the sentiment of the posted content, an approach to correlate the bot accounts to Spanish political parties is proposed, as we will see in the next sections.

III. Research methodology

This section illustrates the steps followed during the development of the research work. In short, this methodology is a natural evolution of the Big Data Mining for Social Bot Identification (BASTION) framework defined in [18] for social bots identification. It also evolves towards an OSINT perspective, integrating the stages defined in [14] for the acquisition of open-source information. Figure 3 shows the followed methodology in this article, defining the modules.

As observed in the figure, the framework is composed of three main components. Specifically, the Data Collection component focuses on the acquisition of Twitter interactions using the Social Feed Manager's crawler and harvester [19]. In turn, collected data goes directly to the Data Analysis component, in which the modules pick relevant features and elicit the users' identification tokens. The resulting data are stored in the Augmented Dataset for further uses. The core of the framework is the Knowledge Extraction component. In here, two machine learning pipelines subsequently attempt to identify and extract information from the collected data. To be more precise, firstly, a supervised ensemble algorithm classifies the Twitter accounts as either humans or bots and, secondly, a series of unsupervised techniques aim to discover correlations between the bots.

A. Data Collection

During the initial Information Discovery, we selected a list of 46 hashtags related to the Spanish general election. More specifically, we gathered a set of hashtags related to significant Spanish political events, as well as one equally distributed among the five main political parties running for the 2019 national election. The complete list is available in the research work's official repository [20]. Notice that we only collected tweets containing at least one of these hashtags.

To gather the relevant content, we deployed the Social Feed Manager (SFM) platform that continuously queried the Twitter API to accumulate the tweets (both original, retweets, replies, and quotes) containing at least one keyword within the sets mentioned above. The collection started on October 4$^{\text{th}}$, 2019, and concluded on November 11$^{\text{th}}$, 2019. During the collection window, the harvester collected around six million tweets and almost a million unique users. Despite the considerable size of the harvested data, we cannot guarantee its completeness due to limitations of the Twitter's standard search APIs. Aiming at reducing this potential gap, we recursively added all those tweets (within the observation period) that were referenced by the collected ones.

To better deal with the unstructured tweet data and to circumvent the relatively fixed structure of the SFM database technology, we exported the data to a non-relational MongoDB instance. At the end of the Data Collection phase, the framework had two complementary collections of data, namely the tweets ($\mathbb{T}$) and the users ($\mathbb{U}$). To be precise:

- The tweets' collection $\mathbb{T}$ stored a JSON document for each tweet containing all the objects provided by the Twitter APIs, e.g., the unique identifier of the tweet, its text or the author, among others. For a full list, see both Twitter API and SFM documentation.
- The users' collection $\mathbb{U}$ stored the set of unique users collected while crawling the tweets' collection.

Throughout the following sections, we will refer to the records in these collections as named tuples and access their fields using the superscript notation. Furthermore, we will use bold $\mathbf{t} \in \mathbb{T}$ and $\mathbf{u} \in \mathbb{U}$ to indicate any tweet $\mathbf{t}$ or user $\mathbf{u}$, respectively. For example, $\mathbf{t}^{\text{uid}}$ and $\mathbf{u}^{\text{uid}}$ indicate the unique identifier of the tweet and the user, while $\mathbf{t}^{\text{text}}$ and $\mathbf{t}^{\text{timestamp}}$ refer to the tweet's text and timestamp.

B. Data Analysis

The Data Analysis component takes care of transforming the raw data obtained by the Twitter API into a usable format to power the machine learning pipelines. This component retrieves the full data from the harvester and, after an anonymization layer, outputs the Augmented Dataset that will be used in the analysis.

Unveiling it, this component hosts both the feature extraction and the anonymization processes. Concerning the former, human analysts are in charge of analyzing the information retrieved from the social media and of discovering, through careful literature review, which features are required by the analysis.

1) Augmented Dataset: The first and foremost augmented feature used by the framework is the tweet's sentiment score. That is, a machine learning classifier predicts a score for each tweets' text, obtaining a value ranging from zero (extremely negative) to one (extremely positive), i.e., $\mathbf{t}^{\text{sent}} \in [0, 1]$. To be precise, the tweet's text also passes through a method that makes the text easier to understand (e.g., removes special characters and converts emoji to text). Retweets in particular will have the same sentiment score as the original tweet.

The second augmented feature used by the framework is the tweet's topic mention. The information and Feature Discovery phases picked up five different groups of keywords ($\mathbb{W}^{\Gamma}$) that refer to an equal number of political events and trending topics, namely: i) election, ii) exhumation, iii) Catalonia, iv) debate and v) AbascalEH. The Feature Model determines, for each bag-of-word $\mathbb{W}^{\Gamma}$, whether at least one of the contained

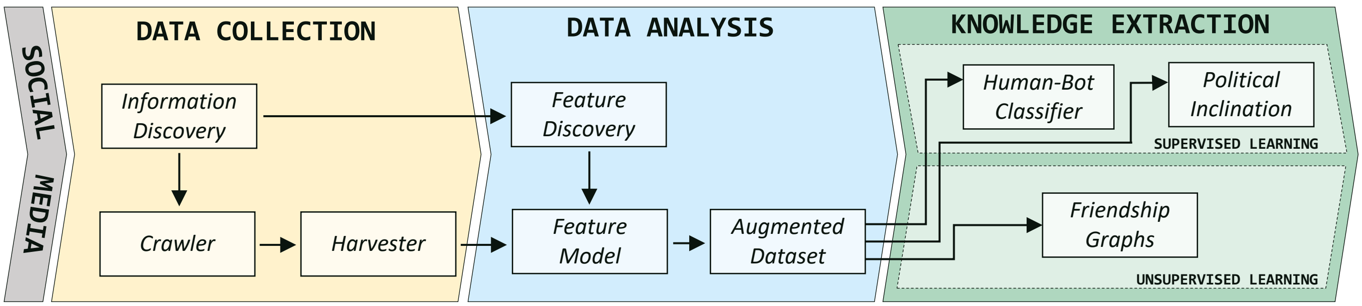


Fig. 3: Research methodology adopted for social bots identification and profiling.

keywords matches in the tweet's text (or original text in case of retweets).

Similarly to the previous group of features, the framework extracts whether the tweets are mentioning any political party. To be precise, the information and discovery phases identified five groups of keywords ($\mathbb{W}^P$, where $P \in \mathbb{P} = \{\mathsf{UP}, \mathsf{PSOE}, \mathsf{Cs}, \mathsf{PP}, \mathsf{VOX}\}$) that refer to the five major political parties in Spain. However, rather than having a generic match with these bag-of-words, these mentions to the political parties are considered exclusive. Formally, each text is labeled in a one-hot-encoding-like style, i.e., a tweet labeled as PP only features at least one mention to the People's Party and, in the text, there is not any match with any keyword included in another political parties' bag-of-word. The same applies for the retweets.

Moreover, the framework extracts all the relations between any two users. In other words, we cave the intersections between any given user's followers and followings lists and the whole list of harvested ones. Formally, let $\mathbb{F}_e$ and $\mathbb{F}_i$ be the followers and followings lists of the user $\mathbf{u}$ provided by the Twitter's API. Then, we define $\mathbf{u}^{\text{fwe}} = \mathbb{F}_e \cap \mathbb{U}$ and $\mathbf{u}^{\text{fwi}} = \mathbb{F}_i \cap \mathbb{U}$ as the user's followers and followings lists respectively.

Finally, the resulting feature set is released from any reference to the original tweets; moreover, each tweet's unique identifier is replaced with a randomly generated universal unique identifier (UUID). These changes are performed automatically, and the map between the tweets' identifiers and the newly generated UUIDs has not been saved to guarantee the unique directionality of the transformation. Note that UUIDs do not replace the users' identifiers at this point.

After the Data Collection and Analysis stages, it was also needed to process the data to acquire useful information about all the accounts identified, being able to classify them as bots or humans, to identify the political inclination of the detected bots, and to identify relationships between bots.

C. *Knowledge Extraction - Supervised Learning*

1) Human-Bot Classifier: To detect the presence of social bots within the dataset, we used Botometer, the *de-facto* standard in terms of bot identification in Twitter [7]. This framework, developed at Indiana University, consists of a machine learning platform that extracts and analyzes more than 1200 features spanning from content-related information, user profile data, and sentiment analysis to produce a score suggesting the likelihood that the account is indeed a social bot. Botometer is publicly available through RapidAPI.

Among the several scores provided by the tool, theoretical analysis and experimental results showed that the most suitable one is the Universal Score $\mathbf{u}^{\text{s}} \in [0, 1]$, as described in [7]. It rates any account according to the likelihood of being a bot, i.e., scores close to 0 indicate that the users are extremely likely to be real-user accounts, while scores nearby 1 show that the accounts are behaving like social bots.

To be as conservative as possible, the Human-Bot Classifier of our proposed framework (see Figure 4) only considers the two ends of the universal score's scale for classification. That is to say, once we obtained the universal score distribution for the Augmented Dataset, a statistical approach was used to label the accounts. Formally, we identified with the 75^{th} percentile ($p_{75} \in [0, 1]$) and 95^{th} percentile ($p_{95} \in [0, 1]$) the boundaries for the human-like and the bot-like classes (establishing a percentile of the 95^{th} is common in the social sciences for statistical significance or detection of outliers [21]). That is to say, we considered as `human` ($\mathbf{h} \in \mathbb{H} \subseteq \mathbb{U}$) those users $\mathbf{u}$ with $\mathbf{u}^{\text{s}} < p_{75}$ and `bots` ($\mathbf{b} \in \mathbb{B} \subseteq \mathbb{U}$) those users $\mathbf{u}$ with $\mathbf{u}^{\text{s}} > p_{95}$, labeling all the users in between as `unclear`. In our sample, with $p_{75} = 0.236$ and $p_{95} = 0.691$, we labeled approximately six hundred thousand accounts as humans, 150 thousand as unclear and forty thousand accounts as social bots.

Finally, the module takes care of saving this information in the dataset and, at last, anonymizes the users' identifier, converting them into randomly generated UUIDs.

2) Political Inclination classifier: In order to be able to solve the attribution problem, as described in [18], the framework includes a political affiliation classifier that aims to sort the social bots according to their political behavior. As reported in [22], [23], natural language processing (NLP) has been proved effective in analyzing the political characteristics of Twitter users. Thus, the Political Inclination classifier makes use of the average sentiment score for each group of political party's keywords ($\mathbb{W}^P$).

Formally, the sentiment score $s_{\mathbf{u},\tau,P} \in [0, 1]$ for a given user $\mathbf{u}$, a tweet type τ and a political party $P \in \mathbb{P}$ is computed as follows:

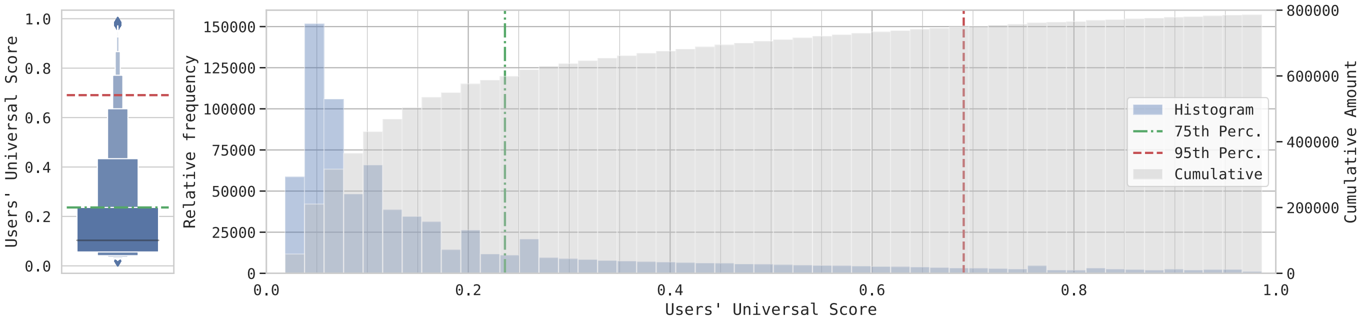


Fig. 4: Users' bot score histogram, with cumulative distribution and percentiles.

$$
\begin{aligned}
s_{\mathbf{u},\tau,P} = \mathtt{mean}(&\mathbf{t}^{\text{sent}} \in [0,1] \mid \mathbf{t}^{\text{uid}} = \mathbf{u}^{\text{uid}} \\
&\wedge \mathbf{t}^{\text{type}} = \tau \\
&\wedge \exists\ w \in \mathbb{W}^{P}\ .\ w \subseteq \mathbf{t}^{\text{text}} \\
&\wedge \forall\ P' \neq P\ .\ \forall\ w' \subseteq \mathbb{W}^{P'}\ .\ w' \nsubseteq \mathbf{t}^{\text{text}})
\end{aligned}
$$

It is noteworthy that, in order to calculate $s_{\mathbf{u},\tau,P}$, we only consider those tweets mentioning a single political party P, that is to say, a tweet citing two or more parties is excluded to avoid user's opinion misinterpretation. Moreover, such sentiment $s_{\mathbf{u},\tau,P}$ is divided according to the tweet type τ to remark the differences in the type of interaction.

Furthermore, for each one of the identified subjects $\mathbb{W}^{\Gamma}$, this module extracts the average sentiment score depending on both the user, the tweet type, and the mentioned political party. In the same way as the previous equation, tweets with multiple political parties' mentions are not included in the analysis. However, this condition does not apply to the subject mentions, that is to say, a tweet is considered if it includes at least, but not limited to, one keyword of that subject. Thus, formally, given a user $\mathbf{u}$, a tweet type τ, a political party $P \in \mathbb{P}$, and a subject $\gamma \in \Gamma$, the sentiment score $s_{\mathbf{u},\tau,P,\gamma} \in [0,1]$ is computed as follows:

$$
\begin{aligned}
s_{\mathbf{u},\tau,P,\gamma} = \mathtt{mean}(&\mathbf{t}^{\text{sent}} \in [0,1] \mid \mathbf{t}^{\text{uid}} = \mathbf{u}^{\text{uid}} \\
&\wedge \mathbf{t}^{\text{type}} = \tau \\
&\wedge \exists\ w \in \mathbb{W}^{P}\ .\ w \subseteq \mathbf{t}^{\text{text}} \\
&\wedge \forall\ P' \neq P\ .\ \forall\ w' \subseteq \mathbb{W}^{P'}.\ w' \nsubseteq \mathbf{t}^{\text{text}} \\
&\wedge \exists\ q \in \mathbb{W}^{\gamma}\ .\ q \subseteq \mathbf{t}^{\text{text}})
\end{aligned}
$$

Thus, combining the equations mentioned above, we obtain a feature vector $\mathbf{x}^{\mathbf{u}}$ for each user $\mathbf{u}$ that includes both the average sentiment score towards any given political party $s_{\mathbf{u},\tau,P}$ and toward any subject thematic in combination with any political party $s_{\mathbf{u},\tau,P,\gamma}$. Formally:

$$
\begin{aligned}
\mathbf{x}^{\mathbf{u}} =& \{ s_{\mathbf{u},\tau,P} \quad \mid \forall\ \tau \in \Pi\ .\ \forall\ P \in \mathbb{P} \} \\
&\cup \{ s_{\mathbf{u},\tau,P,\gamma} \mid \forall\ \tau \in \Pi\ .\ \forall\ P \in \mathbb{P}\ .\ \forall\ \gamma \in \Gamma \}
\end{aligned}
$$

Regarding the training and testing dataset used for building the machine learning classifier, a total of a thousand among the leading and most important verified politicians have been collected and manually labeled by identifying those accounts that were either verified by Twitter or explicitly mentioned the affiliation with a political party in their name or description. Such handcrafted dataset of users provided the training sample of the classifier ($\mathbb{X}$).

As we did not have specific requirements for adopting a certain kind of classification algorithms, we selected six of the most common ones ($f \in \mathbb{F}$) to perform the analysis, namely:

- Random Forest ($f_1 = \text{rf}(\mathbb{X})$) – 10 trees, with minimum leaf size of 5
- Multilayer perceptron - NN ($f_2 = \text{nn}(\mathbb{X})$) – with a single hidden layer, 100 nodes, ReLu activation function, Adam solver with $\alpha = 0.001$, and 200 replicable training interactions
- Support Vector Machine - SVM ($f_3 = \text{svm}(\mathbb{X})$) – $C = 1.0$, $\epsilon = 0.1$ with RBF kernel, and 100 interaction limits with 0.001 numerical tollerance
- Naive Bayes ($f_4 = \text{nb}(\mathbb{X})$)
- k-Nearest Neighbor - kNN ($f_5 = \text{knn}(\mathbb{X})$) – with 5 neighbors, euclidean metric, and uniform weights
- AdaBoost (with internal tree) ($f_6 = \text{ab}(\mathbb{X})$) – with 50 estimators, SAMME.R classification algorithm and linear regression loss function

The evaluation of the algorithms has been carried out with a 10-fold cross validation whose results are available in Table I.

Model	Accuracy	Precision	Recall	F1	AUC
RF	**0.962**	**0.963**	**0.962**	**0.962**	0.995
NN	0.955	0.955	0.955	0.955	0.993
SVM	0.932	0.934	0.932	0.932	0.991
NB	**0.962**	**0.963**	**0.962**	**0.962**	**0.998**
kNN	0.940	0.943	0.940	0.940	0.977
AB	0.954	0.955	0.954	0.954	0.994

TABLE I: Evaluation of the trained classifiers with a manually labeled sample.

We indicate with $f(\mathbf{x}) = [\hat{\mathbf{y}}, p(\hat{\mathbf{y}})]$ the result of applying a classifier f to any given feature vector $\mathbf{x}$, resulting in a set of predicted classes $\hat{\mathbf{y}}$ with their relative probability $p(\hat{\mathbf{y}})$. It holds that the probability of any given class $\mathbf{y'}$ is given by $f(\mathbf{x})[\mathbf{y'}] = p(\mathbf{y'})$. If the evaluated user has a manually verified label, this will be indicated with $\mathbf{y}$.

Given the high scores achieved by all the algorithms, we decided to combine their predictions (weighted accordingly) to label the social bot accounts. Formally, for a party $P \in \mathbb{P}$

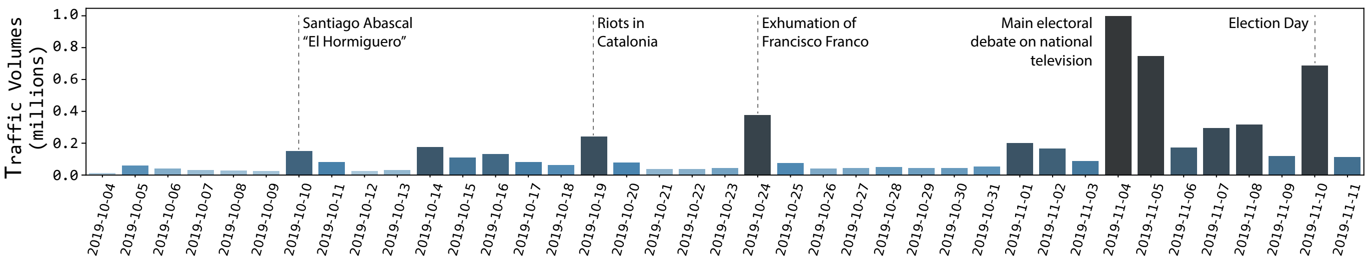


Fig. 5: Tweets volumes per day.

and a user $\mathbf{u}$ with feature vector $\mathbf{x^u}$, the probability of being associated to that political party, $\mathbf{u} \mapsto P$, is expressed as:

$$p(\mathbf{u} \mapsto P) = \underset{\forall\ f \in \mathbb{F}}{\text{mean}}\Big(f(\mathbf{x^u})[P]\Big) \tag{1}$$

In other words, the probability of the predicted label is assigned by averaging the probability assigned by each algorithm. We map an account with a political party if and only if $p(\mathbf{u} \mapsto P) > \delta$, where $\delta \in [0, 1]$ represents a threshold defined as:

$$\delta = 1 - \frac{1}{|\mathbb{P}|} = 0.8 \tag{2}$$

However, if the predicted probability for a single party is lower than δ, then we consider the cumulative sum of the two parties with the highest probability. For example, consider a bot $\mathbf{b} \in \mathbb{B}$ that has $p(\mathbf{b} \mapsto \textsf{PSOE}) = 0.5$ and $p(\mathbf{b} \mapsto \textsf{UP}) = 0.45$. Individually, neither PSOE nor UP reach a good enough confidence to justify the classification. However, when considered together:

$$p(\mathbf{b} \mapsto \{\textsf{PSOE}, \textsf{UP}\}) = p(\mathbf{b} \mapsto \textsf{PSOE}) + p(\mathbf{b} \mapsto \textsf{UP}) = 0.95$$

Then, their cumulative sum is greater than the threshold, hence the predicted class "PSOE-UP" can be accepted.

Finally, we reject the predicted class for any bot for which it is not possible to predict a class with confidence score higher than δ.

D. *Knowledge Extraction - Unsupervised Learning*

The mission of the unsupervised machine learning component is to help identify, even visually, groups of social bots by analyzing their properties as a social group. Indeed, one of the most critical challenges [18] lodges in the visualization of these large graphs. This research makes use of the ForceAtlas2 algorithm [24] that permits to plot up to ten thousand nodes.

To be precise, the framework creates a direct graph using the followers ($\mathbf{u}^{\text{fwe}}$) and followings ($\mathbf{u}^{\text{fwi}}$) lists of harvested users. The generated graph is thus preprocessed to identify and retain only the main component of the network. The resulting users' network is then drawn a painted accordingly to the users' assigned political party, as we will see in Section IV-D.

IV. Experiments and Results

This section introduces the experiments and the results obtained by analyzing the 2019 Spanish general election. All these experiments were conducted in a dedicated server featuring 2 Intel Xeon E5-2630 v4 CPUs (a total of 20 cores at 2.2 GHz) and 80 GB of DDR4 memory at 2400 MHz. The whole project occupies around 115 GB of storage.

A. *Statistical information*

This section presents some descriptive statistics regarding the dataset size and composition.

Tweet Type	Amount	Proportion	User's AVG
Retweet	5,377,150	88.12%	6.83
Original	604,660	9.91%	3.78
Reply	68,365	1.12%	3.25
Quote	52,004	0.85%	2.29
Total	**6,102,179**	100%	-

TABLE II: Tweet types distributions of collected tweets.

Firstly, Table II reports the overall amounts of collected tweets according to their type. That is to say, over the 6,102,179 tweets collected, the vast majority (88.12%) are retweets, followed by originals (9.91%), replies (1.12%), and quotes (0.85%). The table also includes the average (AVG) number of interactions per user.

User Group	Amount	Proportion	User's AVG
Removed	21,914	2.71%	-
Humans	598,898	74.10%	6.20
Uncertain	146,664	18.15%	-
Social bots	40,733	5.04%	4.26
Total	**808,209**	**100%**	-

TABLE III: Collected users' groups distributions.

Secondly, from the tweets we extrapolated the set of users (808,209 unique accounts to be precise) as indicated in Table III; however, a number of accounts (2.71%) have been excluded *a-priori* either because they had been removed, they had no tweets, or they had a private profile. As firstly illustrated in Figure 4 and then presented in Table III, a total of 598,898 have been classified as *Humans* (i.e., $\mathbf{u}^{\text{s}} < p_{75}$), 146,664 as *Uncertain* (i.e., $p_{75} \leq \mathbf{u}^{\text{s}} \leq p_{95}$), and 40,733 as *Bots* (i.e., $\mathbf{u}^{\text{s}} > p_{95}$). To reduce the bias in our investigations, the *Uncertain* group has been discarded.

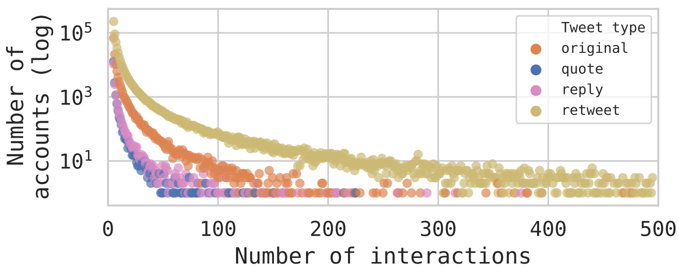
Thirdly, as shown in Figure 5, the collected tweets are not uniformly distributed across the observation period. In the figure, we highlighted several significant political events that are associated with these traffic spikes.

Finally, the traffic volumes, their relative proportions, and the average number of interactions per user are presented in Table IV according to both the users' class and the tweet types.

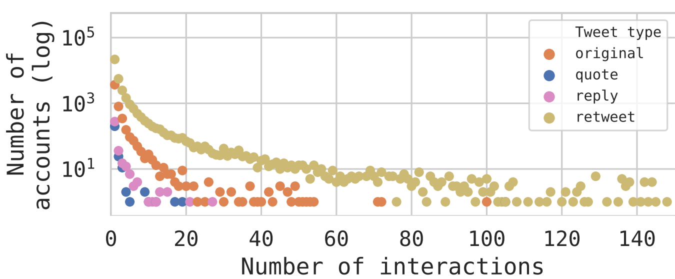
Users' Class	Tweet Type	Amount	Proportion	User's AVG
Humans	Retweet	3,852,066	86.57%	6.98
	Original	498,251	11.20%	3.84
	Reply	55,375	1.24%	3.28
	Quote	43,718	0.98%	2.30
Subtotal		**4,449,410**	**100%**	**6.20**
Social bots	Retweet	168,440	92.54%	4.59
	Original	12,449	6.84%	2.30
	Reply	661	0.36%	1.82
	Quote	475	0.26%	1.94
Subtotal		**182,025**	**100%**	**4.26**
Excluded (missing user)		1,470,717	-	-

TABLE IV: Distribution of tweets per category.

In addition to these high-level metrics, we analyzed the distributions of the users according to their generated traffic volumes. In other words, Figure 6 illustrates the number of users as a function of their number of generated interactions for those accounts classified as humans $\mathbb{H}$ (Figure 6a) and social bots $\mathbb{B}$ (Figure 6b). To improve the figure's readability, the vertical axes feature a logarithmic scale while the horizontal ones are capped both for humans and social bots.



(a) Distributions for those accounts classified as humans.



(b) Distributions for those accounts classified as social bots.

Fig. 6: Number of accounts having a certain amount of interactions, grouped according to the tweet types

As shown in both Figure 6a and Figure 6b, the interactions' volumes do not present a uniform distribution; on the contrary, the vast majority of the users (both human and social bots) have published just a few tweets. Interestingly enough, all tweet types present both similar shares and distributions (although scaled), as numerically reported in Table IV.

B. *Behavioral differences between humans and social bots*

One of the main questions raised by this research is to measure the effectiveness of social bots. A first attempt includes the analysis of the interactions for those tweets that have a direct and unique target, i.e., retweets, replies and quotes. These tweet types indicate that a user is actively interacting with a target on different degrees. For example, a user might retweet content because of the target's idea, or comment on it (either reply or quote) due to shared interests or concerns.

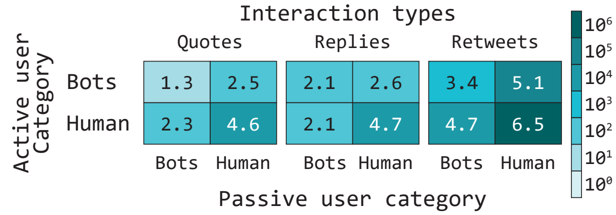


Fig. 7: Interactions involving bots and users by tweet type.

In this context, we denote as "active user" the person who creates the interaction whereas "passive user" is the owner of the retweeted, replied, or quoted tweet. Following this convention, Figure 7 presents the volumes of shared tweets according to both the tweet type (i.e., retweets, replies or quotes) and the users' classification. Note that: *i)* the volumes are presented in their logarithmic form to increase the separability of the smaller categories, and, *ii)* this section only considers those tweets where both the active and the passive users are in the humans or the social bots groups; in other words, tweets starting or targeting an account in the *Uncertain group* (i.e., $p_{75} \leq \mathbf{u}^s \leq p_{95}$) have been excluded (around $500{,}000$ tweets).

Nevertheless, and as expected, the single most common type of interaction is the retweet, providing alone more than 92% of the interactions. Considering the whole scenario, 5% of the traffic volumes involve a social bot either as an active or passive actor. However, if we exclude the human-to-human retweets, the reader might notice that this proportion skyrockets to an overwhelming 66.4%.

On the one hand, a different picture is depicted by looking deeper into the human's interactions other than the human-to-human retweets. Notably, only 37% of these interactions are targeting social bots, of which 95% are retweets. In other words, humans tend to retweet the content shared by the social bots instead of quoting or replying to them. On the other hand, only 2% of social bots activities are targeting other social bots. In particular, social bots tend to retweet human contents in an attempt to make it viral [18]: according to our data, and if we exclude human-to-human retweets, almost half of the generated traffic (45%) is provided by social bots retweeting humans' contents.

C. *Political party affinity*

This section focuses on the analysis of the contents shared by the social bots in an attempt to solve the attribution problem identified in [18]. To do so, only those social bots $\mathbf{b}$ that have at least one tweet specifically targeting a political party P are considered. In this sense, of the nearly forty thousand social bots identified in Table III, only $18{,}518$ qualifies (45.46%), for a total of $182{,}018$ tweets.

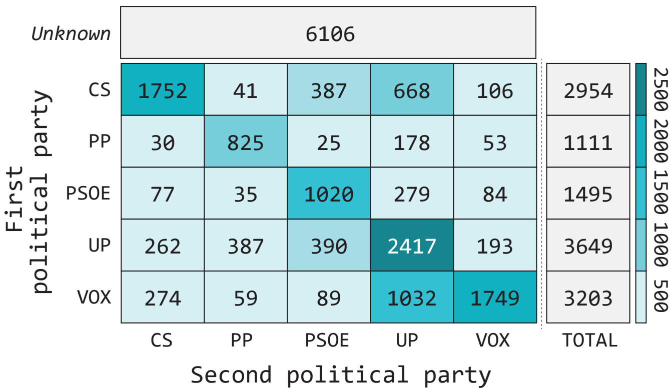


Fig. 8: Number of social bots according to their predicted political party. See Section III for the detailed explanation.

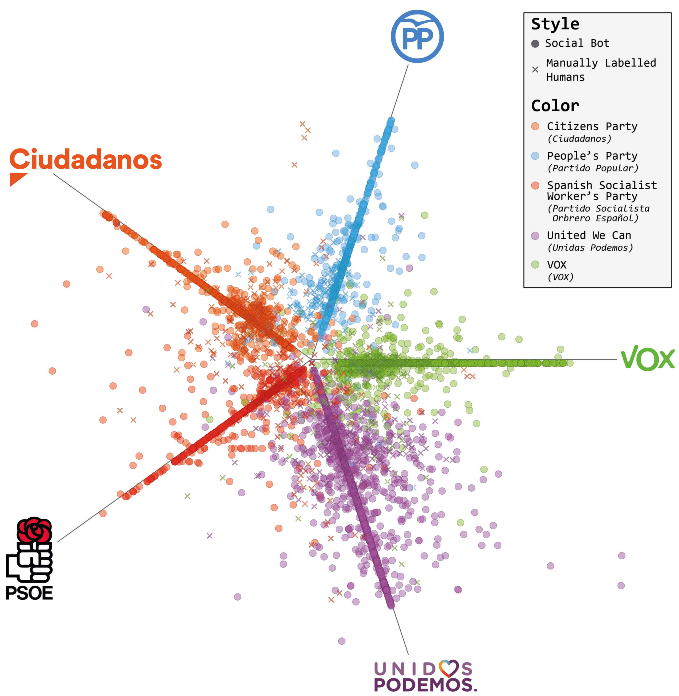


Fig. 9: Bi-dimensional projection of social bots' single political affinities and manually labeled users.

Figure 8 reports the results of the ensemble classifier described in Section IV-C. In the figure, the vertical axis represents the first political party identified by the classifier, while the horizontal presents the second one identified. It follows that the diagonal cells represent those social bots that have been classified as mapped to a single political party. In the figure, those social bots that have not been classified with sufficient enough confidence are labeled as "Unknown".

Classification results of those social bots aligned with only a single party are graphically represented in Figure 9 in conjunction with manually labeled users. To be precise, the figure is the result of a manually-driven dimensionality reduction to five dimensions projected into a plane. Each dimension constitutes the average user's sentiment towards a political party.

From now on, and for the sake of simplicity, we will focus the rest of the analysis on the social bots associated with only one political party.

D. On the subject of social bots, followers, and followings

The groups of social bots identified so far have several common traits. First and foremost is their coordinated behavior. However, along with the shared content, social media like Twitter also feature direct connections between the users (i.e., followers and followings).

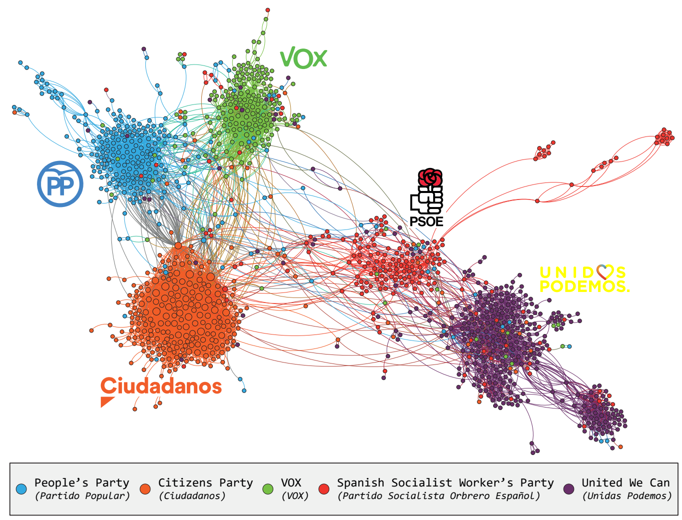


Fig. 10: Friendship relationship among bots.

As a result, Figure 10 presents the social bots as classified in Section IV-C. The figure maps each user with all the accounts that share either a following or a follower relation. The higher the number of connected nodes, the closer the nodes are painted in the plane. Within each cluster, bigger nodes indicate a higher centrality measure of the account, i.e., social hubs. Note that the algorithm used for charting the graph does not consider the political affinities, but only metrics such as the average degree and the centrality values. Thus, the graph has been colored according to the political affinities during the post-production phase. In combination with the quantitative measures reported in the previous sections, Figure 10 further suggests a coordinated effort to deploy, maintain, and employ social bots.

E. The temporal appearance of social bots

The presence of social bots on Twitter on the eve of the 2019 Spanish general election was not random. According to our data, a direct correlation manifests between their appearance, the spikes in the traffic volumes, and the major political events.

To begin with Figure 11, it illustrates the daily appearance of new social bots. On the horizontal axis we have the dates in the observation period, while, on the vertical one, we observe the number of detected social bots. There are two complementary groups of time series in Figure 11. The time series represented by the lines reports the cumulative number of unique social bots, while the histogram ones illustrate the daily activity per political party.

From the figure, it is possible to notice that the increase in the numbers of the social armies was constant throughout the

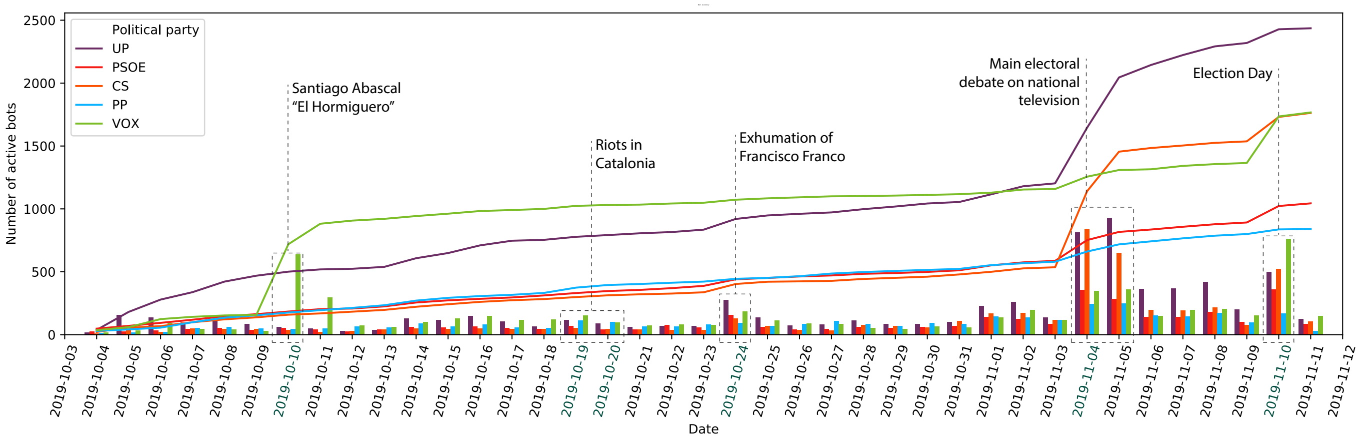
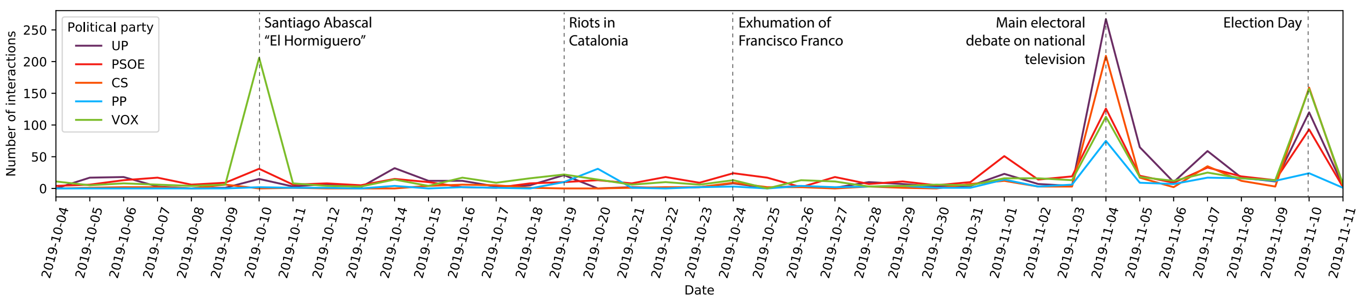
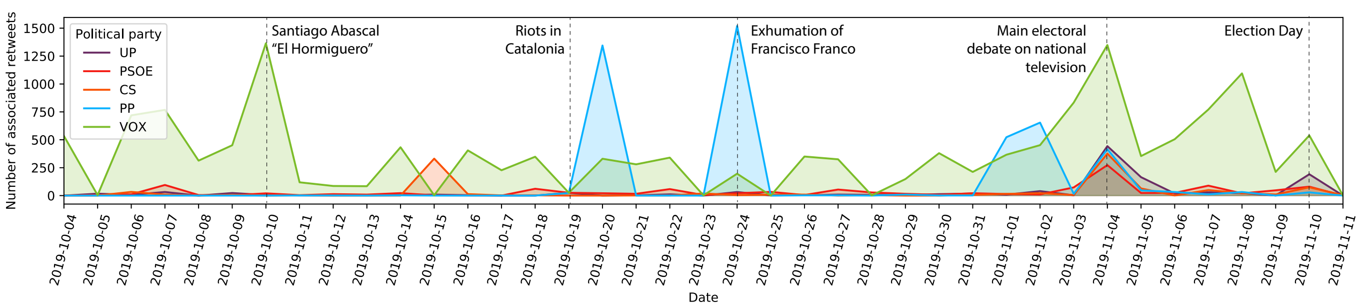


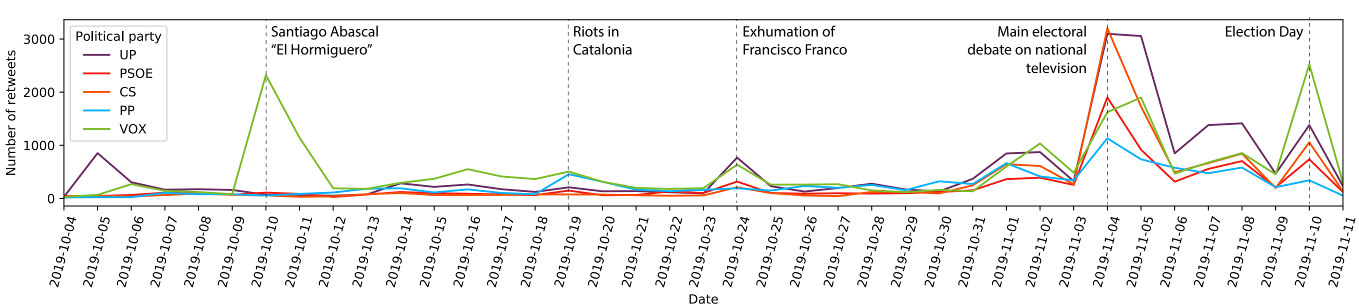
Fig. 11: Active social bots with one-party affinity (cumulative and per day basis)



(a) Originals, quotes and replies.



(b) Associated retweets.



(c) Total retweets.

Fig. 12: Activity of social bots with one-party affinity.

whole electoral season. Notably, more than 500 bots (exposing an affinity with VOX) emerged on the same day as VOX's leader participated in the national TV show "El Hormiguero". Concerning the dates when the riots in Catalonia occurred, our data do not include any significant increase in the number of social bots. However, we did detect a sizable spike during the controversial act of Franco's exhumation.

Notably, during the ten days before the election day, more than five thousand new social bots popped up, mainly during the night of the electoral debate and the election day. During this specific time window, two different patterns appear. On one side, the social bots associated with the PP, PSOE, and VOX parties presented essentially the same growing trend, while those bound to both UP and Cs depict a remarkable increment. One might infer that those fake accounts were precisely hired to support the final events of the campaign.

Although correlation does not imply causality, it is difficult not to suspect that the social bots were driven by actors who followed, monitored, and participated in the Spanish political life. The following sections will attempt to corroborate this hypothesis by looking at the behavior of these social armies.

F. Timeline of the social bots' interactions

The daily increase in the number of social armies in principle does not imply their effectiveness. While, on the one hand, Section IV-B presented the humans-bots interactions, on the other hand, Figure 12 shows their temporal properties.

To be more precise, Figure 12a reports, according to the political party, the number of generated original tweets, quotes and replies created by social bots. As mentioned in the previous sections, all the anomalies in the traffic volumes' patterns are correlated to major political events. As for the retweets, we devote two complementary figures. Figure 12b quantifies the total retweets caused by the social bots contents, whereas Figure 12c reports the volumes of the actual retweets shared directly by the social bots.

In other words, Figure 12b presents the total volume of tweets that retweeted social bots original tweets, quotes and replies. Note that the overall numbers are in the orders of tens of thousands of retweets. By looking at Figure 12b, it appears that the social bots allegedly connected to VOX promote contents that are attractive enough to be shared and propagated by other users. Oddly enough, the tweets originated from social bots affiliated with the PP obtained several thousand retweets on three separate occasions, namely during the riots in Catalonia, during the exhumation of Francisco Franco, and in the days before the national debate. Ostensibly, these social bots shared contents aligned with the political agenda of PP.

Finally, in Figure 12c, it appears that, apart from a few precise exceptions, the amount of retweets during the observation period is consistently below a thousand tweets per group. Besides this steady behavior, further research is required to analyze these retweeting strategies and their interactions with other parties.

G. Tweets' contents analysis

The social bots aspects discussed so far are mostly measures of the account's properties. However, the correlation between their actions and the subjects and ideas promoted is also of great importance.

In this regard, Figure 13 reports the sentiment analysis of each group of social bots in comparison with the manually labeled sample of verified human users. The boxplots in the figure report the distribution of the sentiment score toward each political party, i.e., their attitude concerning the alliances. Both humans (in Figure 13a) and social bots (in Figure 13b) are divided according to their assigned political party. Scores neighboring zero indicate an antagonistic opinion, while a score close to one suggests an appraisal for the subject.

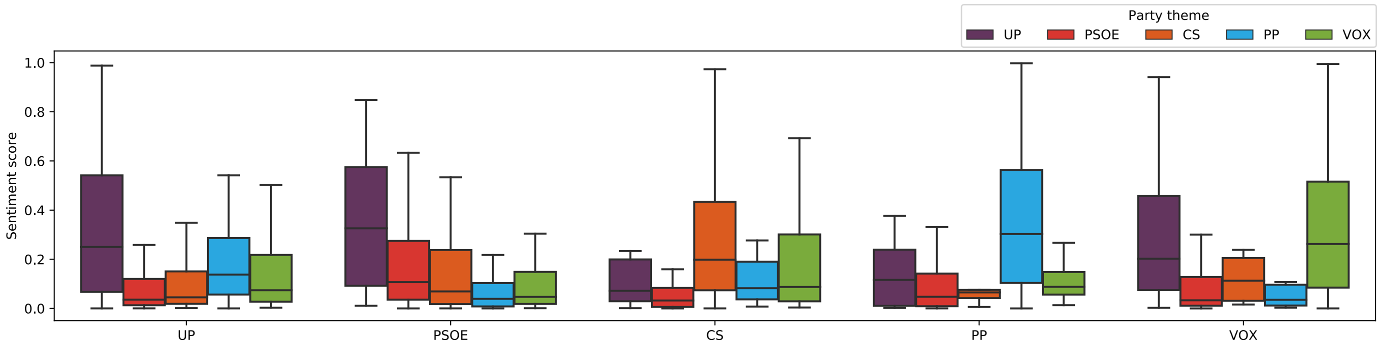
Although the overall perspective around the parties is negative (global averages are below 0.2), each group presents a reasonable change regarding their corresponding party. The political scenario surrounding the Spanish general election might justify the anomalies for some specific combinations. For example, in Figure 13a, the manually verified humans representing the PSOE were substantially pushing more supportive content towards the UP than their own party.

As for the social bots of Figure 13b, the panorama looks appreciably different. As stated before, despite the negatively-oriented sentiment scores that characterize the political discussion, it seems that the ensemble classifier managed to separate the social bots accordingly to the supported political party.

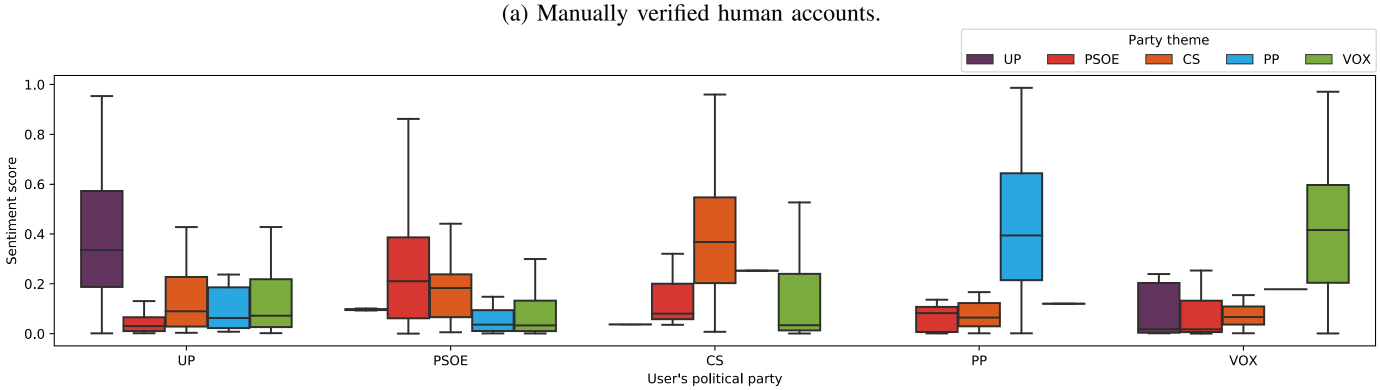
V. Discussion

First, we would like to map the results and data that we observe with the Spanish political context at the time. This election took place due to a previous failed attempt where there was no majority in the parliament, and the PSOE leader did not accomplish to receive support from the rest of the parties to become Prime Minister. Hence, during the pre-electoral period, the political parties were aware of the inevitable necessity to find alliances in other parties as an absolute majority was unlikely to be reached with just the citizens' voting. Figure 13 showed that the overall sentiment trend in our data sample is negative, and this is aligned with the trend in many social networks and communities, where trolling and hate speech are the norm [25]; however, we did report some differences in terms of sentiments as part of the interactions between some parties. For example, in Figure 13a we see a positive sentiment score of PSOE towards UP, and not in vain, the outcome of this election was a coalition government between these two parties. Interestingly enough, the network representation presented in Figure 10 reflects the actual positioning of the political parties. To be more precise, both right-wing (PP and VOX) and left-wing (PSOE and UP) parties appear to be closely connected, using the central party (Cs) as a hub. Indeed, the Citizens party (Cs) reports the highest average degree and centrality values of the network.

Regarding the levels of activity of these accounts in Twitter, perhaps the most noteworthy finding is the very clear association between real-world events with peaks of activity of the different parties; this association between bursts of activity and media events have been further explored by previous work [26]. Some of these events include the visit of Santiago Abascal to "El Hormiguero" associated with a very high peak



(a) Manually verified human accounts.



(b) Social bots.

Fig. 13: Sentiment of labeled users and social bots against party themes.

of activity of VOX, the riots in Catalonia which are associated with activity peaks of both VOX and PP, or the exhumation of Francisco Franco, with a high number of original tweets from PP, and retweets in the case of VOX and UP. Then, on other events, such as the main electoral debate and the elections' day, the political bots of all parties were highly active in terms of interaction volumes. All of these represent key events where political parties can emphasize their views in order to gain votes, and thus this can explain the increased levels of activity at those times. These differences in how the groups of bots of each party have been orchestrated based on different events might be indicative of centralized coordination behind the scenes with clear political goals, perhaps in the form of botnets that should be further studied.

Furthermore, the study also has some limitations that we would like to acknowledge. The first one is that we are using the external tool Botometer to detect these bots, and thus our work can only be as trustworthy as this tool; however, Botometer is widely considered as the best option in the state of the art. Several decisions have been taken based on statistical and empirical measures, and our decisions lie more on the conservative side; therefore, we believe that the real number of bots and influence is significantly higher than the estimates that we report. Some uncertainty is common under this kind of computational studies in social networks where the ground truth is not directly observable.

Finally, we believe this line of work to be of high importance, and the rationale is grounded on psychological research on belief and behavior. Previous researches have shown how social media might be changing the political beliefs of citizens [27], and there have been researches attempting to model how this belief influence might propagate across social networks [28]. A number of theories have connected the beliefs and behavior of people, the latest one in [29] known as the "Reasoned action approach," provides a framework that connects beliefs, intentions, and behavior. Therefore, these social bots, by importantly amplifying and propagating specific ideas, can affect the belief of the social media users, thus directly affecting their behavior when voting in political elections. The majority of social networks including Twitter, currently are loose enough to allow certain automatization. However, given the high-stakes that are at play when these features are misused with ill objectives, we consider vital to keep studying the potential effects of these phenomena on our modern democracies.

VI. Conclusions and Future work

This paper has analyzed the presence and behavior of social bots on Twitter in the 2019 Spanish general election scenario. To summarize the article at hand, the analysis has been performed by following a research methodology composed of three main stages that encompass both supervised and unsupervised learning, namely: data collection, data analysis, and knowledge extraction. As a result, the proposed framework presents capabilities such as human and social bots classification, social bots' political inclinations identification, and a hint of social botnet discovering through friendship analysis.

A pool of experiments has demonstrated not only a non-negligible number of social bots on Twitter participating in the Spanish elections but also a relevant number of daily interactions and traffic volume. Indeed, the analysis of behavioral differences between humans and social bots have detected that humans tend to retweet content shared by the social bots, while social bots tend to retweet human contents to make it viral. Our analysis also analyzed and reported quantitative measurements of the social bots' temporal appearance and relationships. Last but not least, although the sentiment analysis reported an overall negative trend—a common aspect in social media nowadays—it also suggested essential differences between the political parties.

As future work, we plan to extend the harvested dataset with further interactions that would improve the performance of the classifier algorithms, perhaps including OSINT metadata regarding political actors. We also consider the revision of the sentiment analysis process by adding new metrics and improving the classification capabilities, which, to date, are limited for the Spanish language. Finally, we will investigate the presence of potential botnets in the Spanish general election scenario as well as measure their influence in human decisions and the result of the election.

Acknowledgments

This study was partially funded by a grant from the Spanish National Cybersecurity Institute (INCIBE) with code INCIBEI-2015-27353, by the Spanish Government grants FPU18/00304, FJCI-2017-34926 and RYC-2015-18210, co-funded by the European Social Fund, by a predoctoral grant from the University of Murcia and by the Irish Research Council, under the government of Ireland post-doc fellowship (grant code GOIPD/2018/466). Authors would also like to acknowledge Prof. Karl Aberer at EPFL, Prof. Albert Blarer at Armasuisse and IMDEA Networks Institute for their support to this work.

References

[1] E. Zhuravskaya, M. Petrova, R. Enikolopov, Political effects of the internet and social media, SSRN Electronic Journal (2019) 1–30 doi:10.2139/ssrn.3439957.

[2] F. Gómez Mármol, M. Gil Pérez, G. Martínez Pérez, Reporting offensive content in social networks: Toward a reputation-based assessment approach, IEEE Internet Computing 18 (2) (2014) 32–40. doi:10.1109/MIC.2013.132.

[3] S. Bradshaw, P. N. Howard, The global disinformation order: 2019 global inventory of organised social media manipulation, Tech. rep., University of Oxford (2019).

[4] E. Kušen, M. Strembeck, You talkin' to me? Exploring human/bot communication patterns during riot events, Information Processing & Management 57 (1) (2020) 102126. doi:10.1016/j.ipm.2019.102126.

[5] E. Ferrara, The history of digital spam, Communications of the ACM 62 (8) (2019) 82–91. doi:10.1145/3299768.

[6] L. Luceri, A. Deb, S. Giordano, E. Ferrara, Evolution of bot and human behavior during elections, First Monday 24 (9) (aug 2019). doi:10.5210/fm.v24i9.10213.

[7] K. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, F. Menczer, Arming the public with artificial intelligence to counter social bots, Human Behavior and Emerging Technologies 1 (1) (2019) 48–61. doi:10.1002/hbe2.115.

[8] E. Ferrara, Bots, elections, and social media: A brief overview (2020). arXiv:1910.01720.

[9] M. Stella, M. Cristoforetti, M. De Domenico, Influence of augmented humans in online interactions during voting events, PLoS ONE 14 (5) (2019) 1–16. doi:10.1371/journal.pone.0214210.

[10] C. Besel, J. Echeverria, S. Zhou, Full cycle analysis of a large-scale botnet attack on Twitter, in: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2018, pp. 170–177. doi:10.1109/ASONAM.2018.8508708.

[11] A. Badawy, A. Addawood, K. Lerman, E. Ferrara, Characterizing the 2016 russian IRA influence campaign, Social Network Analysis and Mining 9 (1) (2019) 31. doi:10.1007/s13278-019-0578-6.

[12] F. B. Keller, D. Schoch, S. Stier, J. Yang, Political astroturfing on Twitter: How to coordinate a disinformation campaign, Political Communication (2019) 1–25 doi:10.1080/10584609.2019.1661888.

[13] S. Aral, D. Eckles, Protecting elections from social media manipulation, Science 365 (6456) (2019) 858–861. doi:10.1126/science.aaw8243.

[14] J. Pastor-Galindo, P. Nespoli, F. Gómez Mármol, G. Martínez Pérez, The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends, IEEE Access 8 (2020) 10282–10304. doi:10.1109/ACCESS.2020.2965257.

[15] R. Schuchard, A. T. Crooks, A. Stefanidis, A. Croitoru, Bot stamina: Examining the influence and staying power of bots in online social networks, Applied Network Science 4 (1) (2019). doi:10.1007/s41109-019-0164-x.

[16] S. Cresci, M. Petrocchi, A. Spognardi, S. Tognazzi, Better safe than sorry: An adversarial approach to improve social bot detection, in: 10th ACM Conference on Web Science, 2019, pp. 47–56. doi:10.1145/3292522.3326030.

[17] L. Luceri, S. Giordano, E. Ferrara, Don't feed the troll: Detecting troll behavior via inverse reinforcement learning (2020). arXiv:2001.10570.

[18] M. Zago, P. Nespoli, D. Papamartzivanos, M. Gil Pérez, F. Gómez Mármol, G. Kambourakis, G. Martínez Pérez, Screening out social bots interference: Are there any silver bullets?, IEEE Communications Magazine 19 (August) (2019) 2–8. doi:10.1109/MCOM.2019.1800520.

[19] Social feed manager (2016). doi:10.5281/zenodo.597278.

[20] J. Pastor-Galindo, M. Zago, P. Nespoli, S. López Bernal, A. Huertas Celdrán, M. Gil Pérez, J. A. Ruipérez-Valiente, G. Martínez Pérez, F. Gómez Mármol, Botbusters - Analysis of the 2019 Spanish General Election (Mar 2020). doi:10.5281/zenodo.3733195.

[21] B. Finlay, A. Agresti, Statistical methods for the social sciences, Dellen, 1986.

[22] J. A. Caetano, H. S. Lima, M. F. Santos, H. T. Marques-Neto, Using sentiment analysis to define Twitter political users' classes and their homophily during the 2016 American presidential election, Journal of Internet Services and Applications 9 (1) (2018) 18. doi:10.1186/s13174-018-0089-0.

[23] Y. Halberstam, B. Knight, Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter, Journal of Public Economics 143 (2016) 73–88. doi:10.1016/j.jpubeco.2016.08.011.

[24] M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software, PLoS ONE 9 (6) (2014) 1–12. doi:10.1371/journal.pone.0098679.

[25] A. Jakubowicz, et al., Alt_right white lite: Trolling, hate speech and cyber racism on social media, Cosmopolitan Civil Societies: An Interdisciplinary Journal 9 (3) (2017) 41.

[26] Y.-R. Lin, B. Keegan, D. Margolin, D. Lazer, Rising tides or rising stars? dynamics of shared attention on Twitter during media events, PLoS ONE 9 (5) (2014). doi:10.1371/journal.pone.0094093.

[27] N. Swigger, The online citizen: Is social media changing citizens' beliefs about democratic values?, Political Behavior 35 (3) (2013) 589–603. doi:10.1007/s11109-012-9208-y.

[28] H. Nguyen, R. Zheng, Influence spread in large-scale social networks – a belief propagation approach, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, 2012, pp. 515–530. doi:10.1007/978-3-642-33486-3_33.

[29] M. Fishbein, I. Ajzen, Predicting and changing behavior: The reasoned action approach, Taylor & Francis, 2011.